EMERGENT TRADING BIASES IN LLM AGENT SWARMS: BEHAVIORAL-FINANCE BENCHMARKS AND INFORMATION-CHANNEL ABLATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Large Language Model (LLM) agents are being widely adopted to simulate stock market behavior, yet beyond aggregate price dynamics we have limited mechanistic understanding of how individual agents trade when interacting in a multi-agent market: do typical behavioral economic biases show up, and if so, what causes them? We answer both questions with a swarm of 100 LLM agents (Qwen2.5-7B-Instruct) interacting in a simulated market with realistic microstructure. With no bias explicitly mentioned in any prompt, agents are given heterogeneous trading personalities and connected over seeded Barabási–Albert social networks. We measure loss aversion, the disposition effect, herding, anchoring, overconfidence, and panic against a budget-constrained zero-intelligence (ZI) baseline, and test whether the biases documented in human traders emerge. A five-condition ablation (full, no_social, no_personal, no_news, no_persona) across seven stress events and a no-event control (20 seeds, 800 runs) measures how much each information channel influences each bias, with all comparisons seed-clustered and FDR-corrected. Four of the six biases emerge in the expected human direction: herding at 15.4$\times$ the ZI floor, the disposition effect—on a single-asset adaptation of Odean's PGR–PLR construction—within the range of human traders (0.034 vs. 0.050, a directional comparison rather than a level match), loss aversion higher than the floor (market-drift controlled, 1.49 vs. 1.13), and panic elevated above its own no-event floor in 18 of 20 seeds. Conversely, anchoring arises with a reversed sign, and overconfidence is absent in trade sizing. Some causes are more predictable, such as loss aversion eroding when the personal position layer is ablated (1.94 to 1.38), as prospect theory anticipates. Others are more difficult to predict: removing persona heterogeneity nearly triples herding, and panic, while amplified by news, is diminished by social context. No ablation leads to perfectly rational agents, and a residue persists under every prompt we test. Understanding each information channel's influence, as well as that residue, tells us how bias can be controlled when deploying multi-agent systems.

1 INTRODUCTION

When a swarm of LLM agents trade in a shared market, affecting one another through prices and the actions their neighbours can see, the resultant market looks remarkably like a real market: volatility clustering, fat-tailed returns, bubbles, and even crashes are reproduced (Yang et al., 2025; Zhang et al., 2024). These simulations are already being used to test economic hypotheses, with the agents acting as replacements for human participants (Horton, 2023). Outside of simulations, LLMs are also being increasingly trusted with real financial decisions (Financial Stability Board, 2024). However, before these use cases scale, we need to better understand how these agents behave. Do they act rationally or do human biases creep into their decisions?

Observing solely the swarm's aggregate behavior does not reveal whether its agents are rational. A market can look realistic without containing a single realistic trader. In fact, it is known that random traders on a budget still produce efficient allocations (Gode & Sunder, 1993) and that simple mechanical rules can produce phenomena like fat tails (Lux & Marchesi, 1999; Cont & Bouchaud, 2000). What we do know is that LLMs tested in isolation exhibit typical human-like biases. Whether

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

those biases survive, compound, or diminish once many agents interact in a market is unknown. We explore this with a focus on the behavior of individual agents: do the biases observed in human traders appear in a swarm, and if so, what causes them?

This paper systematically answers both questions. In our swarm, no prompt mentions or paraphrases any cognitive bias (one disclosed exception, Swarm Setup). Current frameworks such as TwinMarket (Yang et al., 2025) aim for price realism by equipping agents with empirically calibrated bias parameters — our goal is to study the opposite, so we withhold all bias content and observe what happens on its own. All emergent biases can then be attributed to the model's own decision-making. We also simulate controls to make that emergence measurable. A zero-intelligence (ZI) baseline trades the identical market, helping us understand the bias level produced by market mechanics (Gode & Sunder, 1993), and an ablation grid one-by-one removes information channels, so that each bias can be traced back to the context layer that causes it, or to the model itself.

We ask three primary questions, and answer each.

- **RQ1: Do typical human biases emerge when no prompt mentions them?** Four of the six we focus on do, in the expected human direction. The disposition effect lands within the human range (0.034 vs. Odean's 0.050), loss aversion is elevated above the ZI floor under drift control (1.49 vs. 1.13), herding is measured at 15.4$\times$ the floor, and panic is elevated above its own no-event floor in 18 of 20 seeds. The other two do not: overconfidence does not appear, and anchoring is measured but sign-reversed. However, knowing that the biases emerge is only half the picture. To control them, we need to understand what causes them.
- **RQ2: What is the total effect of each information channel?** Each bias decomposes differently. For example, loss aversion sits at 1.94 with every channel intact, but removing the personal position layer drops it to 1.38 (close to the ZI floor of 1.13), and removing the news layer has no impact. This is intuitive and corroborates prospect theory: reference points matter (Kahneman & Tversky, 1979) and narrative does not. Panic, on the other hand, is amplified by the news layer and diminished by the social layer (1.92 vs. 2.25 in no_social).
- **RQ3: Do the biases emerge because of the prompts or the model itself?** Both, unevenly. We run an ablation with no personas, and notice that loss aversion is almost completely eroded, herding nearly triples, and anchoring flips sign. A residue survives on three of six metrics, so the measured biases split into those carried by the persona prompts and those that do not depend on the prompt at all, and knowing which is which is what makes them controllable.

2 Related Work

Cognitive Biases in Single LLMs. Individual LLMs are known to exhibit certain human biases. They anchor on numbers present in their prompts, and their answers shift predictably when a question is reframed (Jones & Steinhardt, 2022). Fine-tuning a model to follow instructions can amplify its biases, exacerbating decoy and certainty effects (Itzhak et al., 2024). When making decisions including risky payoffs, LLMs exhibit framing-dependent risk preferences consistent with prospect theory (Payne, 2025). LLMs also fall for the same reasoning traps that catch humans, though chat tuning can reduce or even reverse these fallacies (Binz & Schulz, 2023; Hagendorff et al., 2023; Jia et al., 2024; Raman et al., 2024). Many of these are the same biases that behavioral finance measures in human traders. However, these results come from testing single models. Whether the same biases appear in a population of interacting agents has not been tested against behavioral finance benchmarks.

LLM agent societies and market simulations. LLM agents interacting in shared worlds began with small generative agent towns (Park et al., 2023) but have scaled to societies of thousands. They are being used to study information spread, social dynamics, and herd effects on social platforms (Piao et al., 2025; Yang et al., 2024). In classic economic games, their choices often overlap with those of human participants (Horton, 2023; Aher et al., 2023; Mei et al., 2024; Xie et al., 2024). None of this work, however, examines trading. In finance, TwinMarket models a market of LLM investors and reproduces bubbles, crashes, and fat tails (Yang et al., 2025), but its agents are calibrated with empirically measured bias parameters by design, so bias there is an input rather than a

finding. StockAgent and EconAgent build trading and macroeconomic agent architectures (Zhang et al., 2024; Li et al., 2024), and MarS trains a generative order-flow engine (Li et al., 2025); their validation is aggregate realism, which as argued above leaves individual trading behavior unexamined.

Other work studies behavioral questions such as bubble formation and speculation (Lopez-Lira, 2025; Ouyang & Sui, 2026; Saxena et al., 2026; Li et al., 2026), but none of it measures the canonical biases of behavioral finance or asks where they come from. Most similar to our work, Henning et al. (2025) study agents in small experimental-finance groups and find that LLM agents fail to replicate human results; the mechanism behind the failure is left open. Coletta et al. (2024) test whether LLMs can imitate subrational behavior when instructed to, effectively the inverse of our question. Jia et al. (2025) separate intrinsic reasoning from contextual influence across game-theoretic settings. We ask a similar question, but in a swarm of agents interacting in a market. What remains open is the question we began with: do the canonical trading biases emerge without being prompted, and what causes them? We address this gap directly. To our knowledge, no prior system combines prompt-neutrality that keeps bias content out of the agents' context entirely, behavioral-finance estimators read against a seed-matched zero-intelligence floor that separates measured bias from market mechanics, and a channel ablation grid that isolates the contribution of each context layer to each bias.

Biases. Every measure used is taken from behavioral-finance literature. Loss aversion comes from prospect theory (Kahneman & Tversky, 1979), the disposition effect estimator from Odean's brokerage-record analysis (Shefrin & Statman, 1985; Odean, 1998; Frazzini, 2006), and the herding statistic from the LSV measure of trade clustering (Lakonishok et al., 1992; Barber et al., 2009). Overconfidence, anchoring, and panic are grounded in their own empirical literatures (Barber & Odean, 2001; Gervais & Odean, 2001; George & Hwang, 2004; Shiller & Pound, 1989; Hong et al., 2004). These papers provide both the estimators and, where applicable, the human magnitudes to benchmark against.

3 SWARM SETUP

Agents. We use 100 LLM agents (Qwen2.5-7B-Instruct with temperature 0.7) with eight equally distributed trading personas: momentum, value, fundamentalist, noise, risk-averse retail, contrarian, informed, and institutional. Personas describe trading style (the signals they should trade on and how aggressively) in completely symmetrical language, and no prompt names or defines any cognitive bias. We withhold bias *terminology* rather than all strategy content, and the clearest case is the noise trader, which treats the actions of others as tradable information. This mirrors the standard definition of the persona (Black, 1986). However, we see that this exception does not drive the herding results: herding *rises* when the agents cannot see their neighbors, and is at its peak under the no_persona ablation, showing that the herding measured is not an artifact of this persona's prompt. Rather than assuming any implicit bias-inducing content is absent, we measure it via the no_persona ablation, which runs the same grid giving all agents homogeneous personas with a neutral prompt.

At each time step, the agent is given certain context (see below) and returns a JSON action {buy, sell, hold} along with a quantity and stated confidence. The Kelly Criterion at half strength is used to scale position size based on this stated confidence (Kelly, 1956; MacLean et al., 2010). If any agent's response cannot be parsed correctly, it resorts to holding the asset. Any run whose final steps are dominated by these fallbacks, which are characteristic of a server crash, is discarded. However, no such runs were observed in the final grid.

Context and information channels. Every agent's prompt includes four layers: (i) **market**, the current price, bid-ask quote, 5- and 20-time-step rolling returns, rolling volatility, and a summary of any relevant fundamentals, all lagged by one time step such that agents at t see the order book as of $t-1$; (ii) **social**, the most recent actions made by an agent's immediate neighbors; (iii) **personal**, the agent's current position, average entry price, and unrealized P&L, the minimal state required to observe reference-point behavior (Kahneman & Tversky, 1979), with no realized P&L history and no record of past wins or losses; and (iv) **news**, text describing any news events. For ablations no_personal, no_social, and no_news, the respective channels are ablated and the rest included.

Market. Agents trade a single anonymized mid-cap enterprise-software asset against a market maker that always fills orders, with linear price impact ($\lambda = 0.01$), a 10 bp spread that widens during liquidity events, Student-t ($\nu = 4$) exogenous shocks (Praetz, 1972; Cont, 2001), and a weak Ornstein–Uhlenbeck pull toward fundamental value as in ABIDES (Uhlenbeck & Ornstein, 1930; Byrd et al., 2020). Appendix A gives the full specification.

Network. Agents are connected by a Barabási–Albert network with $m = 3$ (Barabási & Albert, 1999), producing a few hubs and many sparsely connected agents, a structure used to model investor information networks (Ozsoylev & Walden, 2011). Modeling other topologies (Watts & Strogatz, 1998; Erdős & Rényi, 1959) is left to follow-on work.

Events. All runs last 200 steps. At step 100, one of seven news events, or the no_event control, is applied to the news context layer. The events used are a Fed surprise, flash crash, VIX spike, earnings miss, short squeeze, sector rotation, and liquidity crisis. After 10 steps the event is removed from the news layer; however, any changes to fundamentals remain (e.g., analyst targets cut after an earnings miss). Event narrative and information therefore decay on different timescales.

4 Experimental Setup

Conditions. We run six conditions on the same seeds and events: 5 LLM conditions $\times$ 8 events $\times$ 20 seeds = 800 runs, plus 160 ZI runs (960 total). This includes a budget-constrained zero-intelligence (ZI) baseline (Gode & Sunder, 1993) where agents randomly pick actions with equal probabilities rather than consulting a persona. The magnitude of bias produced by this condition can be attributed solely to market mechanics, so the results from other conditions are compared to this to check for elevated levels. The other five conditions vary what context the agents have access to. full provides access to all layers, no_personal, no_social, and no_news each remove that respective layer, and no_persona keeps all layers but strips agents of their persona, giving all 100 of them a neutral prompt. A bias that survives a neutral prompt must come from the model's training itself or its alignment tuning (Itzhak et al., 2024; Sharma et al., 2024).

Compute and Reproducibility. The 5 conditions $\times$ 8 events $\times$ 20 seeds $\times$ 100 agents $\times$ 200 time steps result in roughly 16 million LLM calls, and the ZI baseline adds 160 CPU-only runs. At $\sim$13.5 minutes/run, the 800-run grid takes roughly 180 A100-hours (Qwen2.5-7B-Instruct served by vLLM on Google Colab). We will release the specific seeds used, the event library, the persona prompts, and all analysis on publication. Paid APIs or proprietary models were not used.

Statistical protocol. A seed generates a specific combination of network structure, persona assignment and order, and noise. All runs that share a seed face the same world, and since their outcomes are correlated, all statistics are clustered at the seed level. All confidence intervals (CIs) are 95% bootstrap CIs, and are calculated by resampling the 20 seed-level means rather than the 160 individual runs. For hypothesis tests, we find each seed $\times$ condition value by averaging over the 8 events. We then compare full, the ZI floor, and each of the ablations for all 6 metrics. The resulting 30 contrasts are each tested with a paired Wilcoxon signed-rank test across the 20 seeds and jointly BH-FDR-corrected (Benjamini & Hochberg, 1995; Harvey et al., 2016). 19 of the 30 survive and are significant at $q < 0.05$. The no_persona results are also tested against the ZI floor, as this is indicative of the residue. Three of six metrics stay significantly above the floor in the expected direction ($q \leq 0.007$; see RQ3). As a robustness check, the entire grid is recomputed with the realized drift held at zero via regression adjustment. After this adjustment, every bias conclusion remains true.

5 Measuring Bias Emergence

All metrics are computed from executed trades and positions, never from agent self-reports. W and L denote completed round-trip trades closed at a gain or a loss (FIFO accounting). We call a statistic *elevated* if it clears the ZI floor and moves in the direction humans show. The floor is needed because an estimator's rational value, stated for each metric below, is not what it actually returns in practice. ZI agents have no cognition at all, and they still produce LAR 1.13, panic 1.51, H 0.004, and β -0.11. Clearing the floor therefore rules out the estimator itself and the market mechanism. It does

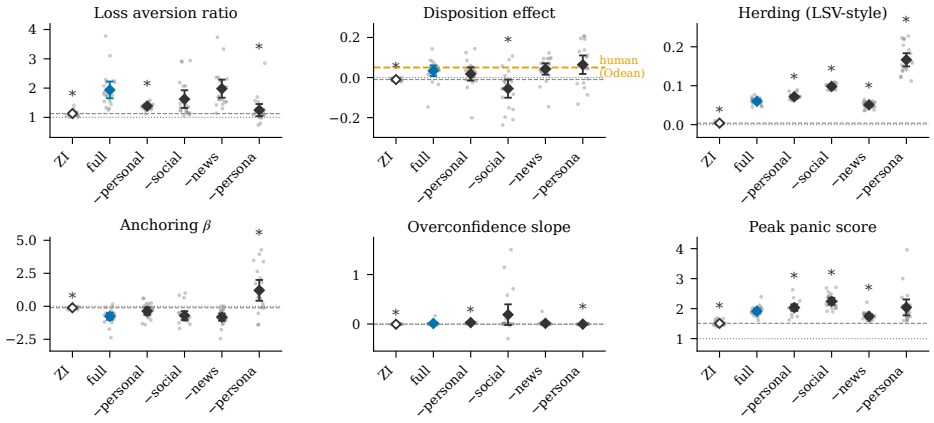

Figure 1: Seed-level results for all six biases under every condition. Gray dots are the 20 per-seed means, diamonds are arm means with seed-clustered 95% CIs, the dashed line is the ZI floor, and the dotted line is the rational baseline. $*$ marks arms that differ from full at BH-FDR $q < 0.05$. The orange line in the DE panel is Odean's human estimate of 0.050.

not rule out a rational trader doing the same thing, so we do not treat elevation alone as evidence of a bias.

Loss aversion ratio (LAR). $\mathrm{LAR} = \bar{\tau}_L/\bar{\tau}_W$, the mean holding time of losers over winners. A reluctance to realize losses holds losers longer. Human data attest to the direction (Odean, 1998) but no holding-time ratio, so LAR is read against the rational baseline of 1.0 and the ZI floor.

Disposition effect (DE). $\mathrm{DE} = \mathrm{PGR} - \mathrm{PLR}$, the proportion of gains realized minus the proportion of losses realized (Odean, 1998). We count every holding step, while Odean counts sale days, so only the direction is comparable. The only metric with a numerical human benchmark, 0.050 for retail and 0.02–0.03 for institutions (Frazzini, 2006). Rational value 0.

LSV herding (H). $H = \overline{|p_t - \bar{p}|} - \mathbb{E}_0|p_t - \bar{p}|$, where p_t is the buy fraction among active agents at step t and $\mathbb{E}_0$ is the expectation under independent trading (Lakonishok et al., 1992). It measures same-side trade clustering beyond chance. Human values are 0.027 for institutions and 0.068–0.128 for individuals (Lakonishok et al., 1992; Barber et al., 2009). However, these cluster trades across stocks rather than over time, so all comparisons made are order-of-magnitude only. Rational value 0.

Anchoring (β_{anchor}). The regression coefficient of an agent's sell decision on the distance between current price and entry price. Positive β means attraction to the anchor (George & Hwang, 2004). No human benchmark, so only the sign is interpreted. Rational value 0.

Overconfidence slope. The slope of position size on the agent's current winning streak (Gervais & Odean, 2001). Since size is a deterministic function of stated confidence (Swarm Setup), the slope tests whether confidence tracks wins. Rational value 0; no human benchmark.

Panic score. Peak sell-order rate relative to the run's first-quarter baseline, measuring flight-to-safety selling (Brunnermeier & Pedersen, 2009). A maximum over 200 noisy steps has a high floor, so panic is tested against the ZI floor and its no-event floor of 1.77, not the rational value of 1.

6 Results

Figure 1 shows every arm at seed resolution. Table 1 gives exact values. Figure 2 shows event dynamics and the mechanism.

RQ1: Which biases emerge? All six tests survive against the ZI floor after seed-level correction. LAR, herding, and panic are unanimous across all 20 seeds ($q = 8.2 \times 10^{-6}$), anchoring follows at $q = 9.1 \times 10^{-5}$, DE and overconfidence at $q = 2.7 \times 10^{-3}$. Differing from the floor only means

Table 1: Seed-level condition means for the six primary biases. * marks contrasts that differ from full at seed-clustered paired-Wilcoxon BH-FDR $q < 0.05$ (19 of 30). Standard deviations and the full condition × event grid are in Appendix B.

	LAR	DE	H	β	OC	Panic
ZI floor	1.13*	−0.010*	0.004*	−0.11*	−0.005*	1.51*
full	1.94	0.034	0.060	−0.76	0.011	1.92
no_personal	1.38*	0.018	0.072*	−0.36	0.029*	2.03*
no_social	1.62	−0.055*	0.098*	−0.70	0.188	2.25*
no_news	1.98	0.043	0.051*	−0.82	0.014	1.75*
no_persona	1.25*	0.064	0.167*	1.21*	−0.003*	2.04

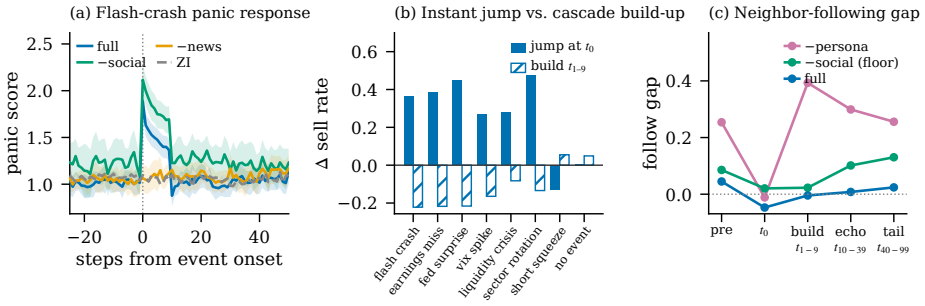

Figure 2: (a) Panic around the flash crash by condition (mean ± 95% CI over seeds). (b) Sell-rate jump at $\tau = 0$ and build over t_{1-9} per event, full condition. (c) Neighbor-following gap Δ by event window, sell-side events pooled.

the bias cannot be explained by market mechanics. Only four of the six also differ from the rational baseline in the expected human direction.

Loss aversion best illustrates how we separate real bias from one caused by mechanics. Agents hold losers 1.94× longer than winners (CI $[1.70, 2.23]$) against a ZI floor of 1.13. However, some of that gap is caused by market mechanics and not purely by the LLMs' decision-making. Over a 200-step run with no event, the full market ends 38.97% lower, while the seed-matched ZI population ends 10.36% higher. This decline is itself an interesting result, and we see that the swarm buys 0.45 shares for every one it sells, against 1.01 for ZI. Removing news does not affect this (-39.42%), removing social reduces but does not eliminate it (-29.38%), but withholding the position layer completely erodes it (-2.62%), as does a neutral prompt (-2.80%). The depth of the fall predicts over half of a run's LAR (R^2 0.54 per run, 0.63 per seed). Intuitively, agents in a falling market hold losers for longer. Matching full and ZI half-runs of equal drift gives LAR 1.49 vs. 1.13 (seed-clustered $p = 3.4 \times 10^{-4}$), elevated in every bin. Where drift is almost absent, LAR holds at 1.47 while ZI collapses to 1.04. These comparisons require overlapping drift, so we report them alongside the raw 1.94.

The disposition effect offers the best comparison to human data due to the availability of Odean's metrics. His estimator yields 0.050 for real retail investors, 0.02–0.03 for institutions, and -0.010 for the ZI floor. In our experiment, his estimator yields 0.034 (95% CI $[0.008, 0.056]$), putting it within the human range. We tested each individual world separately and found a significant effect in 13 of the 20 worlds. Ten of these demonstrate clear disposition behavior, three show a significant reversal, and the remaining seven show neither (these per-world tests pool all five conditions and are descriptive; they sit outside our FDR-corrected protocol). A variance decomposition attributes 23% of DE variance to the seed and a further 32% to the seed-by-condition interaction, against 19% residual and 1% for which event fired, so what makes DE vary is primarily which world an agent population lands in. Finally, we applied the same drift matching as done for LAR and found that DE *rises* to 0.051 vs. -0.011 (seed-clustered $p = 1.9 \times 10^{-5}$), so the price crash actually makes our value conservative.

Herding showcases the largest elevation: 0.060 versus a floor of 0.0039, a 15.4× ratio on an identical estimator. It also occurs at the same order of magnitude as herding measured in human retail trading (0.068–0.128) (Barber et al., 2009), although the cited statistic clusters trades across stocks per month while ours clusters them over time on one asset. Panic peaks at 1.92 vs. 1.51, and clears its no-event floor of 1.77 in 18 of 20 seeds ($p = 1.9 \times 10^{-5}$).

The remaining two biases do not emerge as they do in humans. Overconfidence differs from the ZI floor (0.011 vs. -0.005) but not from rational behavior. Its slope is indistinguishable from 0 (CI $[-0.007, 0.029]$), so it does not emerge in sizing. Meanwhile, anchoring emerges with a reversed sign from what is expected (-0.76 vs. ZI -0.11), informing us that agents trade *against* their reference level rather than toward it. The persona ablation reverses this (RQ3).

RQ2: Layer attribution. We ablate one context layer at a time and measure the impact on each bias. The results vary by bias.

Loss aversion, as expected, is caused by the personal position layer. Removing this layer causes LAR to collapse from 1.94 to 1.38 ($q = 7.3 \times 10^{-5}$), a substantial move towards the floor. Removing news has no impact (1.98), and removing the social layer causes a drop (1.62) that does not survive correction. Thus, among the context layers, only the entry price affects loss aversion, not the news or neighbors.

Herding moves under all ablations. It falls without news (0.060 to 0.051) and rises without the personal (0.072) or social (0.098) layer. These movements can be explained by the same underlying principle: when agents receive inputs that are more similar, they are more likely to herd. In this case, news is the one input that is common to all agents, so its presence makes contexts more alike and its removal makes them less so. The personal and social layers, on the other hand, are the main differentiators between agents, so removing them has the opposite effect. This phenomenon is most visible when personas are removed entirely, something we will analyze further in RQ3. Herding and panic are also the only two biases that respond to the news channel at all.

Panic has two causes with opposite signs. News amplifies it (1.92 with, 1.75 without), while the social layer dampens it (1.92 with, 2.25 without). Effectively, agents that can see their neighbors tend to panic less. One fact to keep in mind is that panic is generally elevated under no_social. It peaks at 2.04 even during runs with no event at all, so a part of the 2.25 here reflects this underlying mechanism rather than event-driven panic. However, Figure 2a better shows the response with baselines in view and confirms the dampening.

The **disposition effect** and **anchoring** show little movement across ablations. DE has one large change, occurring with the removal of the social layer (a sign flip to -0.055), that vanishes under drift adjustment ($+0.024$). We read this as an artifact of diverging price paths and not a causal social effect. Anchoring shows no significant movement at all, under any layer ablation. For both, attribution is settled by the persona ablation (RQ3). For completeness, panic and the overconfidence slope also rise in the no_personal ablation, movements we did not predict.

Lastly, our events split into two kinds. Only the earnings miss changes the asset's fundamentals, and the rest are pure coverage. Removing the news layer makes the story events vanish completely, and their sell rates replicate levels from the no_event control. However, the earnings miss has an impact via degraded fundamentals, and only this event leaves a price change that is still visible 90 steps later (-13.5 pp). In short, media coverage moves prices only temporarily, while changes to the underlying fundamentals persist.

Extended analysis: Is herding social? A 15.4× elevation in the herding statistic invites a seemingly obvious hypothesis: agents copy their neighbors. Our experimental design allows us to test this directly, and it is interesting to see that this is not the case.

Firstly, the timing rules out a cascade. If agents were copying each other, selling would grow over the course of a few steps as the behavior spreads from one agent to another. That is not what happens. Selling is observed to peak at $\tau = 0$, the same step the news arrives, and per-step herding jumps from 0.049 to 0.252 at this point. Since agents see their neighbors' actions with a one-step delay, we know that, at this point, they can only see what their neighbors did before the event, so there is nothing to copy yet. Selling does fall after the peak (Figure 2b), but this is not because agents calm down. Those still holding shares keep selling at high rates of 0.73–0.93, and the overall reduction in selling is simply a by-product of fewer agents having any open positions to sell.

Secondly, seeing neighbors sell changes nothing. We measure $\Delta = P(\text{sell} \mid \geq 2 \text{ neighbors sold}) - P(\text{sell} \mid \text{no neighbor sold})$. If agents follow their neighbors, Δ should be positive. In full, Δ never exceeds its value measured in no_social, where agents cannot see their neighbors at all, so any Δ there reflects agents reacting to the same underlying events rather than copying, or herding with, one another (Figure 2c).

Thirdly, we observe that hubs do not amplify the response. If this influence flowed through the social network, the best-connected agents would contribute more strongly to the herding mechanism. This is not the case, and we see that hubs sell at almost the same rate as peripheral agents.

The elevated herding we measure is therefore the 100 agents reacting to the same headline rather than to each other. Importantly, this does not weaken the herding result. The LSV estimator counts correlated trading from any source, and this is equally true when measuring herding in human trader data. Whether correlated trades reflect genuine imitation or a shared reaction to the same information is difficult to disentangle in human markets, because both look identical to the estimator. Our ablation grid lets us tell the two apart. The elevation is also not an artifact of our events. Events carry under 1% of herding's variance, and clustering stays near 0.060 even in runs where no event ever fires. In those calm markets the only input agents share is the price itself, and shared price information is enough to synchronize their trades.

There is, interestingly, one condition where true following does emerge. Under no_persona, Δ turns genuinely positive, reaching $+0.39$ in the steps immediately following the event, exactly where copying first becomes possible. Agents whose neighbors are selling become more likely to sell, a true instance of copying. This positive control is what makes the null in full meaningful. Cascades are evidently possible in this architecture, so their absence is a property of diverse agents, not of our experimental design. When the agents are made identical, cascades emerge on their own. These window statistics are descriptive.

RQ3: Prompt versus model. Removing the personas splits the biases into two families, role-conditioned (carried by persona descriptions) and prompt-invariant (survive with no persona at all). Loss aversion belongs to the first family. In the no_persona ablation, it falls to 1.25, telling us that 85% of the full-vs-ZI gap was caused by the personas. The remaining 15% is not significant ($q = 0.26$), so we claim no LAR residue. DE, herding, and panic belong to the second family, and they remain significantly above the ZI floor even under the neutral prompts ($q \leq 0.007$). Anchoring also clears the floor, but with a reversed sign, so we do not count it as persisting. Since no prompt change removes the other three, we label them the residue. Herding rises from 0.060 to 0.167 once all agents share a neutral prompt. This continues the pattern from RQ2, where herding was proportional to how similar agents' inputs were; having the same persona and prompt is essentially as similar as inputs can get. One notable case is the short squeeze event, where all 100 agents buy the moment the news arrives (buy rate 1.000), causing the price to rise by 35% before falling back, effectively replicating a bubble. Apart from this, persona homogeneity also leads to real copying, as mentioned and measured in the previous paragraph. Both of these results support warnings about model monoculture, where a high percentage of market participants rely on the same foundation model. We elaborate on this in the Discussion section.

Anchoring encounters the largest change out of any bias. It flips sign, from -0.76 under full to $+1.21$ under no_persona, and the flip is consistent across most worlds (full negative in 19 of 20 seeds, no_persona positive in 15 of 20). Anchoring under homogeneous agents matches what is expected by the literature, and it is interesting that the addition of persona descriptions inverts this behavior. DE and panic change the least, and are not significantly different from full.

7 Discussion

Human biases are seen to emerge in LLM trading swarms. This emergence is selective — some biases appear only when certain information is present, others do not appear at all, and each can be traced back to its cause. For experimental design, the consequence is that bias is now something that can be engineered, rather than just accepted. A builder who wants to mitigate loss aversion in their agents can withhold the personal position display, or weight it differently. One who wants less herding should make sure to preserve persona diversity. However, what cannot be tuned away is the residue, which survived every prompt change we made. A swarm reproduces some human

biases and does not reproduce others, and which ones depends on its exact configuration. Any study that leans on a particular bias should therefore first check the level at which the bias emerges in its own setup, comparing against behavioral benchmarks rather than stylized facts (Cont, 2001), and a simple ZI floor with a no-event arm makes that check cheap.

To consolidate per-bias results and make swarms easy to compare at a glance, a logical next step is a unified metric of how human-like a swarm's trading is. This would be useful when testing agents before deployment, or for comparing models in a controlled manner. Our results implicitly sketch the requirements for this metric, which include measurement floors, world-level variation, and estimator coupling. Ultimately, we leave its construction to future work.

Model-intrinsic residual bias. Where does the surviving bias come from? The evidence points towards the model's training data itself, and the residue appears to be inherited. Observing residual bias under a neutral prompt is consistent with the idea of models internalizing biased decision language from human text (Payne, 2025; Itzhak et al., 2024). Effectively, the model has learned how humans trade, including the parts that humans typically get wrong. As a result, assuming LLMs to be perfectly rational agents would be incorrect, a consequence of the way they are fundamentally trained. Whether this inheritance varies across models or tuning styles remains untested, and an experiment that runs a cross-model comparison is a strong direction for future work.

Homogeneity as systemic risk. The herding explosion under identical prompts has a simple explanation. Since every agent is drawn from the same model, under the same prompt they react to the same input in the exact same way. Heterogeneous personas dampen this by making agents read the same input differently, the behavioral analogue of heterogeneous-agent stabilization in classical market models (Hommes, 2006). The genuine copying is a separate effect. Since it only appears when personas are absent, it reinforces the idea that personas act as decision policies strong enough to override neighbors' signals. Regulators have long warned that if many market participants start to rely on the same few foundation models, their trading will become highly correlated (Gensler & Bailey, 2020; Financial Stability Board, 2024). Until now that warning was theoretical, but our ablation allows us to observe the result experimentally. As the adoption of LLM trading agents grows and clusters around a handful of models, ensuring the diversity of agents built on top of them requires prioritization. The effects caused by homogeneity are not ones we want to encounter in a real market of deployed agents.

8 LIMITATIONS

All experiments use a single model, so our results are specific to Qwen2.5-7B-Instruct and are not representative of LLMs in general. Temperature was also fixed at 0.7. At lower temperatures identical prompts converge by construction, so homogeneity effects could depend on this.

The neutral prompt is a single instantiation, without paraphrase robustness. The market is just a single asset with market-maker execution rather than a full limit-order book (Byrd et al., 2020), so diversification and cross-asset biases are excluded by construction. The 200-step horizon restricts analysis of long-run equilibria. Under persona conditions the no-event market drifts (-39% in full vs. $+10\%$ for ZI), so all event effects are measured against seed-paired no-event controls, and the drift analyses in RQ1 bound its influence. Only DE provides a numerical human benchmark. While herding and loss aversion have order-of-magnitude and directional comparators, the rest are compared to rational and zero-intelligence baselines. With 20 seeds, correlations are detectable only at $|\rho| \gtrsim 0.58$. Lastly, our event narratives are synthetic and occur for a controlled period of time, whereas real news flow is noisier and often endogenous. Most of these constraints are deliberate controls to allow conclusive attribution, and future work involves varying each of these.

9 CONCLUSION

Simulating a market of 100 LLM agents, we observed human trading biases emerge without being written into any prompt. Our ablation study then isolates where they come from. Some of the bias is carried by the context layers and personas, and some by the model itself. The experiment shows that LLM agents act neither as rational traders nor as replicators of human traders. Both the biases and their sources can now be measured, something that makes them controllable.

AI USE STATEMENT

We used an LLM for research ideation, to stress-test hypotheses before the experiments were designed. An LLM also wrote initial structure for parts of the simulation and analysis code, all of which we reviewed, edited and tested. After the paper was written, we used an LLM to critique it and to cut redundant passages. It did not draft the paper or produce any reported number. We take responsibility for the final content of this work.

REFERENCES

Gati Aher, Rosa I. Arriaga, and Adam Tauman Kalai. Using large language models to simulate multiple humans and replicate human subject studies. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, 2023. Oral. arXiv:2208.10264.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286 (5439):509–512, 1999. doi: 10.1126/science.286.5439.509.

Brad M. Barber and Terrance Odean. Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 116(1):261–292, 2001. doi: 10.1162/003355301556400.

Brad M. Barber, Terrance Odean, and Ning Zhu. Systematic noise. *Journal of Financial Markets*, 12(4):547–569, 2009. doi: 10.1016/j.finmar.2009.03.003.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.

Marcel Binz and Eric Schulz. Using cognitive psychology to understand gpt-3. *Proceedings of the National Academy of Sciences*, 120(6):e2218523120, 2023. doi: 10.1073/pnas.2218523120.

Fischer Black. Noise. *The Journal of Finance*, 41(3):528–543, 1986. doi: 10.1111/j.1540-6261.1986.tb04513.x.

Markus K. Brunnermeier and Lasse Heje Pedersen. Market liquidity and funding liquidity. *The Review of Financial Studies*, 22(6):2201–2238, 2009. doi: 10.1093/rfs/hhn098.

David Byrd, Maria Hybinette, and Tucker Hybinette Balch. ABIDES: Towards high-fidelity multi-agent market simulation. In *Proceedings of the 2020 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation (SIGSIM-PADS '20)*, pp. 11–22. ACM, 2020. doi: 10.1145/3384441.3395986.

Andrea Coletta, Kshama Dwarakanath, Penghang Liu, Svitlana Vyetrenko, and Tucker Balch. Llm-driven imitation of subrational behavior: Illusion or reality? *arXiv preprint arXiv:2402.08755*, 2024.

Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001. doi: 10.1088/1469-7688/1/2/304.

Rama Cont and Jean-Philippe Bouchaud. Herd behavior and aggregate fluctuations in financial markets. *Macroeconomic Dynamics*, 4(2):170–196, 2000. doi: 10.1017/S1365100500015029.

Paul Erdős and Alfréd Rényi. On random graphs I. *Publicationes Mathematicae Debrecen*, 6: 290–297, 1959.

Financial Stability Board. The financial stability implications of artificial intelligence. Technical report, Financial Stability Board, 2024. November 14, 2024.

Andrea Frazzini. The disposition effect and underreaction to news. *The Journal of Finance*, 61(4): 2017–2046, 2006. doi: 10.1111/j.1540-6261.2006.00896.x.

Gary Gensler and Lily Bailey. Deep learning and financial stability. *SSRN working paper 3723132*, 2020. November 1, 2020.

Thomas J. George and Chuan-Yang Hwang. The 52-week high and momentum investing. *The Journal of Finance*, 59(5):2145–2176, 2004. doi: 10.1111/j.1540-6261.2004.00695.x.

Simon Gervais and Terrance Odean. Learning to be overconfident. *The Review of Financial Studies*, 14(1):1–27, 2001. doi: 10.1093/rfs/14.1.1.

Dhananjay K. Gode and Shyam Sunder. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, 101(1):119–137, 1993. doi: 10.1086/261868.

Thilo Hagendorff, Sarah Fabi, and Michal Kosinski. Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in chatgpt. *Nature Computational Science*, 3:833–838, 2023. doi: 10.1038/s43588-023-00527-x.

Campbell R. Harvey, Yan Liu, and Heqing Zhu. . . . and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68, 2016. doi: 10.1093/rfs/hhv059.

Thomas Henning, Siddhartha M. Ojha, Ross Spoon, Jiatong Han, and Colin F. Camerer. Llm agents do not replicate human market traders: Evidence from experimental finance. *arXiv preprint arXiv:2502.15800*, 2025. v3, October 2025.

Cars H. Hommes. Heterogeneous agent models in economics and finance. In Leigh Tesfatsion and Kenneth L. Judd (eds.), *Handbook of Computational Economics, Volume 2: Agent-Based Computational Economics*, chapter 23, pp. 1109–1186. Elsevier, Amsterdam, 2006.

Harrison Hong, Jeffrey D. Kubik, and Jeremy C. Stein. Social interaction and stock-market participation. *Journal of Finance*, 59(1):137–163, 2004. doi: 10.1111/j.1540-6261.2004.00629.x.

John J. Horton. Large language models as simulated economic agents: What can we learn from homo silicus? *arXiv preprint arXiv:2301.07543*, 2023.

Itay Itzhak, Gabriel Stanovsky, Nir Rosenfeld, and Yonatan Belinkov. Instructed to bias: Instruction-tuned language models exhibit emergent cognitive bias. *Transactions of the Association for Computational Linguistics*, 12, 2024. arXiv:2308.00225.

Jingru Jia, Zehua Yuan, Junhao Pan, Paul E. McNamara, and Deming Chen. Decision-making behavior evaluation framework for LLMs under uncertain context. In *Advances in Neural Information Processing Systems 37*, pp. 113360–113382, 2024.

Jingru Jia, Zehua Yuan, Junhao Pan, Paul E. McNamara, and Deming Chen. LLM strategic reasoning: Agentic study through behavioral game theory. In *Advances in Neural Information Processing Systems 38*, 2025. arXiv:2502.20432.

Erik Jones and Jacob Steinhardt. Capturing failures of large language models via human cognitive biases. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. arXiv:2202.12299.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979. doi: 10.2307/1914185.

J. L. Kelly, Jr. A new interpretation of information rate. *Bell System Technical Journal*, 35(4): 917–926, 1956. doi: 10.1002/j.1538-7305.1956.tb03809.x.

Josef Lakonishok, Andrei Shleifer, and Robert W. Vishny. The impact of institutional trading on stock prices. *Journal of Financial Economics*, 32(1):23–43, 1992. doi: 10.1016/0304-405X(92)90023-Q.

Junjie Li, Yang Liu, Weiqing Liu, Shikai Fang, Lewen Wang, Chang Xu, and Jiang Bian. Mars: a financial market simulation engine powered by generative foundation model. In *International Conference on Learning Representations (ICLR)*, 2025. arXiv:2409.07486.

Nian Li, Chen Gao, Mingyu Li, Yong Li, and Qingmin Liao. Econagent: Large language model-empowered agents for simulating macroeconomic activities. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2024. arXiv:2310.10436.

Zeping Li, Guancheng Wan, Keyang Chen, Yu Chen, Yiwen Zhao, Philip Torr, Guangnan Ye, Zhenfei Yin, and Hongfeng Chai. Behavioral consistency validation for llm agents: An analysis of trading-style switching through stock-market simulation. *arXiv preprint arXiv:2602.07023*, 2026.

Alejandro Lopez-Lira. Can large language models trade? testing financial theories with llm agents in market simulations. *arXiv preprint arXiv:2504.10789*, 2025.

Thomas Lux and Michele Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500, 1999. doi: 10.1038/17290.

Leonard C. MacLean, Edward O. Thorp, and William T. Ziemba. Long-term capital growth: The good and bad properties of the kelly and fractional kelly capital growth criteria. *Quantitative Finance*, 10(7):681–687, 2010. doi: 10.1080/14697688.2010.506108.

Qiaozhu Mei, Yutong Xie, Walter Yuan, and Matthew O. Jackson. A turing test of whether ai chatbots are behaviorally similar to humans. *Proceedings of the National Academy of Sciences*, 121(9):e2313925121, 2024. doi: 10.1073/pnas.2313925121.

Terrance Odean. Are investors reluctant to realize their losses? *The Journal of Finance*, 53(5): 1775–1798, 1998. doi: 10.1111/0022-1082.00072.

Shumiao Ouyang and Pengfei Sui. Dissecting ai trading: Behavioral finance and market bubbles. *arXiv preprint arXiv:2604.18373*, 2026.

Han N. Ozsoylev and Johan Walden. Asset pricing in large information networks. *Journal of Economic Theory*, 146(6):2252–2280, 2011. doi: 10.1016/j.jet.2011.10.003.

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST)*, 2023. doi: 10.1145/3586183.3606763.

Kenneth Payne. An analysis of ai decision under risk: Prospect theory emerges in large language models. *arXiv preprint arXiv:2508.00902*, 2025.

Jinghua Piao, Yuwei Yan, Jun Zhang, Nian Li, Junbo Yan, Xiaochong Lan, Zhihong Lu, Zhiheng Zheng, Jing Yi Wang, Di Zhou, Chen Gao, Fengli Xu, Fang Zhang, Ke Rong, Jun Su, and Yong Li. Agentsociety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society. *arXiv preprint arXiv:2502.08691*, 2025.

Peter D. Praetz. The distribution of share price changes. *Journal of Business*, 45(1):49–55, 1972.

Narun Krishnamurthi Raman, Taylor Lundy, Samuel Joseph Amouyal, Yoav Levine, Kevin Leyton-Brown, and Moshe Tennenholtz. STEER: Assessing the economic rationality of large language models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *PMLR*, pp. 42026–42047, 2024.

Maxime Saxena, Marco Pangallo, Cars Hommes, Fabio Caccioli, and R. Maria del Rio-Chanona. Machine spirits: Speculation and adaptation of llm agents in asset markets. *arXiv preprint arXiv:2604.18602*, 2026.

Mrinank Sharma, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askell, Samuel R. Bowman, Newton Cheng, Esin Durmus, Zac Hatfield-Dodds, Scott R. Johnston, Shauna Kravec, Timothy Maxwell, Sam McCandlish, Kamal Ndousse, Oliver Rausch, Nicholas Schiefer, Da Yan, Miranda Zhang, and Ethan Perez. Towards understanding sycophancy in language models. In *International Conference on Learning Representations (ICLR)*, 2024. arXiv:2310.13548.

Hersh Shefrin and Meir Statman. The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, 40(3):777–790, 1985. doi: 10.1111/j.1540-6261.1985.tb05002.x.

Robert J. Shiller and John Pound. Survey evidence on diffusion of interest and information among investors. *Journal of Economic Behavior & Organization*, 12(1):47–66, 1989. doi: 10.1016/0167-2681(89)90076-0.

George E. Uhlenbeck and Leonard S. Ornstein. On the theory of the brownian motion. *Physical Review*, 36(5):823–841, 1930. doi: 10.1103/PhysRev.36.823.

Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998. doi: 10.1038/30918.

Chengxing Xie, Canyu Chen, Feiran Jia, Ziyu Ye, Shiyang Lai, Kai Shu, Jindong Gu, Adel Bibi, Ziniu Hu, David Jurgens, James Evans, Philip H. S. Torr, Bernard Ghanem, and Guohao Li. Can large language model agents simulate human trust behavior? In *Advances in Neural Information Processing Systems 37*, pp. 15674–15729, 2024.

Yuzhe Yang, Yifei Zhang, Minghao Wu, Kaidi Zhang, Yunmiao Zhang, Honghai Yu, Yan Hu, and Benyou Wang. Twinmarket: A scalable behavioral and social simulation for financial markets. *arXiv preprint arXiv:2502.01506*, 2025. v5, October 2025.

Ziyi Yang, Zaibin Zhang, Zirui Zheng, Yuxian Jiang, Ziyue Gan, Zhiyu Wang, Zijian Ling, Jinsong Chen, Martz Ma, Bowen Dong, Prateek Gupta, Shuyue Hu, Zhenfei Yin, Guohao Li, Xu Jia, Lijun Wang, Bernard Ghanem, Huchuan Lu, Chaochao Lu, Wanli Ouyang, Yu Qiao, Philip Torr, and Jing Shao. Oasis: Open agent social interaction simulations with one million agents. *arXiv preprint arXiv:2411.11581*, 2024.

Chong Zhang, Xinyi Liu, Zhongmou Zhang, Mingyu Jin, Lingyao Li, Zhenting Wang, Wenyue Hua, Dong Shu, Suiyuan Zhu, Xiaobo Jin, Sujian Li, Mengnan Du, and Yongfeng Zhang. When ai meets finance (stockagent): Large language model-based stock trading in simulated real-world environments. *arXiv preprint arXiv:2407.18957*, 2024. Published in ACM TIST.

A SIMULATOR SPECIFICATION

The market is designed to expose agents to realistic price dynamics and frictions. A single anonymized asset is used to avoid the influence of any company-specific priors. The asset is solely described at the sector level as a mid-cap enterprise-software company. Agents trade with a market maker that always fills orders, and executed net order flow moves the mid-price linearly ($\lambda = 0.01$). The bid-ask spread is standardized at 10 bp, widening only during liquidity events. Prices are also impacted by factors beyond agent trading, and exogenous shocks are sampled from a Student-t distribution ($\nu = 4$) to ensure that large jumps occur at a similar frequency to real return data (Praetz, 1972; Cont, 2001). Finally, a weak Ornstein–Uhlenbeck term (Uhlenbeck & Ornstein, 1930) pulls the price toward the asset's fundamental value, as in ABIDES (Byrd et al., 2020). The pull is strong enough to ensure prices do not diverge over a run, but weak enough to allow the agents' trading to dominate dynamics.

Agents are connected by a Barabási–Albert network (Barabási & Albert, 1999), grown by adding nodes one at a time. Every new node connects to $m = 3$ existing nodes, chosen with a probability proportional to their current degree (preferential attachment). This produces a network with a few hubs and some sparsely connected agents, a structure used to model investor information networks (Ozsoylev & Walden, 2011).

B FULL CONDITION × EVENT GRID

This appendix gives the full condition-by-event grid behind Table 1. For each of the six bias metrics we report the mean and standard deviation across the 20 seeds in every condition/event cell.

Table 2: Loss aversion ratio (LAR): mean ± sd across the 20 seeds, by condition and event.

Condition	no event	Fed surpr.	flash crash	VIX spike	earn. miss	squeeze	sector rot.	liq. crisis
ZI floor	1.13±0.10	1.13±0.10	1.13±0.10	1.13±0.10	1.13±0.10	1.13±0.10	1.13±0.10	1.13±0.10
full	2.00±0.99	1.99±0.58	1.98±0.81	2.02±0.75	2.03±0.62	1.52±0.40	1.96±0.70	2.02±0.80
no_personal	1.38±0.12	1.40±0.11	1.38±0.11	1.36±0.14	1.41±0.15	1.35±0.18	1.38±0.10	1.40±0.11
no_social	1.73±1.01	1.62±0.87	1.48±0.42	1.73±1.02	1.61±0.83	1.46±0.47	1.47±0.51	1.89±1.23
no_news	2.05±0.86	1.74±0.67	2.11±0.96	2.00±0.86	2.06±0.78	1.93±0.65	1.84±0.82	2.10±0.82
no_persona	1.13±0.50	1.31±0.52	1.23±0.51	1.23±0.49	1.22±0.47	1.46±0.42	1.21±0.47	1.19±0.69

Table 3: Disposition effect (DE): mean ± sd across the 20 seeds, by condition and event.

Condition	no event	Fed surpr.	flash crash	VIX spike	earn. miss	squeeze	sector rot.	liq. crisis
ZI floor	-0.009±0.005	-0.009±0.005	-0.009±0.005	-0.009±0.005	-0.009±0.005	-0.009±0.005	-0.009±0.005	-0.018±0.010
full	0.031±0.062	0.027±0.080	0.030±0.074	0.045±0.062	0.033±0.072	0.047±0.058	0.029±0.069	0.031±0.070
no_personal	0.048±0.082	0.010±0.075	0.017±0.077	0.029±0.079	-0.053±0.068	0.059±0.063	0.011±0.080	0.022±0.079
no_social	-0.023±0.095	-0.068±0.115	-0.052±0.116	-0.048±0.102	-0.070±0.127	-0.044±0.113	-0.088±0.102	-0.050±0.105
no_news	0.048±0.053	0.046±0.068	0.045±0.083	0.047±0.068	0.036±0.072	0.049±0.062	0.033±0.072	0.038±0.066
no_persona	0.101±0.134	0.061±0.119	0.057±0.153	0.079±0.141	0.057±0.141	-0.004±0.129	0.058±0.113	0.104±0.128

Table 4: LSV herding (H): mean ± sd across the 20 seeds, by condition and event.

Condition	no event	Fed surpr.	flash crash	VIX spike	earn. miss	squeeze	sector rot.	liq. crisis
ZI floor	0.004±0.002	0.004±0.002	0.004±0.002	0.004±0.002	0.004±0.002	0.004±0.002	0.004±0.002	0.004±0.003
full	0.051±0.012	0.064±0.010	0.061±0.010	0.061±0.011	0.053±0.012	0.064±0.012	0.064±0.012	0.062±0.009
no_personal	0.059±0.010	0.075±0.008	0.076±0.008	0.074±0.007	0.078±0.015	0.067±0.012	0.072±0.006	0.073±0.010
no_social	0.091±0.012	0.103±0.012	0.096±0.009	0.095±0.010	0.097±0.013	0.100±0.014	0.104±0.011	0.100±0.009
no_news	0.052±0.007	0.049±0.010	0.049±0.009	0.052±0.012	0.049±0.014	0.052±0.013	0.051±0.011	0.054±0.010
no_persona	0.163±0.055	0.156±0.037	0.160±0.035	0.162±0.039	0.159±0.033	0.215±0.044	0.154±0.038	0.164±0.050

Table 5: Anchoring (β_{anchor}): mean ± sd across the 20 seeds, by condition and event.

Condition	no event	Fed surpr.	flash crash	VIX spike	earn. miss	squeeze	sector rot.	liq. crisis
ZI floor	-0.10±0.06	-0.10±0.06	-0.10±0.06	-0.10±0.06	-0.10±0.06	-0.10±0.06	-0.10±0.06	-0.19±0.09
full	-0.83±0.63	-0.80±0.62	-0.77±0.80	-0.72±0.65	-0.86±0.64	-0.49±0.84	-0.70±0.71	-0.93±0.64
no_personal	-0.11±0.65	-0.44±0.64	-0.43±0.67	-0.34±0.66	-0.78±0.59	0.06±0.62	-0.45±0.67	-0.42±0.59
no_social	-0.12±0.96	-0.86±0.97	-0.79±0.98	-0.79±0.88	-0.73±1.01	-0.57±1.27	-0.86±0.87	-0.91±0.88
no_news	-0.85±0.67	-0.68±0.67	-0.82±0.64	-0.87±0.69	-0.93±0.71	-0.75±0.74	-0.72±0.75	-0.96±0.57
no_persona	2.04±2.72	1.14±1.98	1.37±2.60	1.54±2.61	1.27±2.56	-0.82±0.78	1.18±1.76	1.93±2.77

Table 6: Overconfidence slope: mean ± sd across the 20 seeds, by condition and event.

Condition	no event	Fed surpr.	flash crash	VIX spike	earn. miss	squeeze	sector rot.	liq. crisis
ZI floor	-0.005±0.003	-0.005±0.003	-0.005±0.003	-0.005±0.003	-0.005±0.003	-0.005±0.003	-0.005±0.003	-0.005±0.003
full	0.016±0.066	0.008±0.028	0.016±0.049	0.019±0.082	0.012±0.051	0.002±0.016	-0.001±0.015	0.012±0.040
no_personal	0.030±0.015	0.028±0.016	0.027±0.014	0.025±0.013	0.041±0.033	0.025±0.014	0.027±0.013	0.026±0.015
no_social	0.055±0.820	0.130±1.074	-0.047±0.382	0.085±0.352	0.055±0.815	0.565±1.630	0.467±1.291	0.197±0.732
no_news	0.018±0.066	0.027±0.119	0.015±0.059	0.010±0.052	0.018±0.086	0.006±0.024	0.003±0.032	0.014±0.049
no_persona	-0.004±0.006	-0.003±0.007	-0.002±0.007	-0.003±0.006	-0.002±0.005	-0.003±0.005	-0.003±0.006	-0.005±0.006

Table 7: Panic score: mean ± sd across the 20 seeds, by condition and event.

Condition	no event	Fed surpr.	flash crash	VIX spike	earn. miss	squeeze	sector rot.	liq. crisis
ZI floor	1.51±0.08	1.51±0.08	1.51±0.08	1.51±0.08	1.51±0.08	1.51±0.08	1.51±0.08	1.50±0.09
full	1.77±0.19	2.13±0.31	1.97±0.22	1.79±0.17	1.96±0.28	1.80±0.20	2.15±0.23	1.77±0.12
no_personal	1.87±0.22	2.30±0.27	2.00±0.30	1.86±0.24	2.11±0.28	1.97±0.26	2.35±0.29	1.83±0.19
no_social	2.04±0.27	2.49±0.40	2.22±0.32	2.14±0.34	2.18±0.20	2.22±0.34	2.56±0.26	2.13±0.27
no_news	1.77±0.16	1.73±0.15	1.74±0.16	1.73±0.20	1.75±0.17	1.74±0.18	1.73±0.18	1.78±0.20
no_persona	2.04±0.62	2.08±0.73	1.99±0.41	2.06±0.62	2.08±0.60	2.05±0.49	2.03±0.64	2.01±0.53